METHOD FOR SCALABLE, FAST NORMALIZATION OF XML DOCUMENTS FOR INSERTION OF DATA INTO A RELATIONAL DATABASE

ABSTRACT

Disclosed is a method of transferring data from a hierarchical file (having a hierarchical structure, e.g., a markup language file) to a relational database structure (made up of columns and rows. Before processing the actual data, the invention first partitions the hierarchical structure into sections, where each section is dedicated to at least one node of the hierarchical structure. The partitioning process is based on the document type definition file, which is separate from, and different than the hierarchical file. After completing the partitioning, the invention then parses the actual data contained in the hierarchical data file to produce a stream of data pairs and end of section indicators. During the data parsing process, the invention loads the data pairs into corresponding "sections" (created prior to the parsing process) as the data pairs are output from the parsing process. The invention also transfers the node data from these sections to the columns and rows of the relational database structure.